

PCT

RAW SEQUENCE LISTINGPATENT APPLICATION: **US/10/018,009A**DATE: 08/16/2004

TIME: 17:10:33

Input Set : N:\FATIMA\10018009.txt
Output Set: N:\CRF4\08162004\J018009A.raw

3 <110> APPLICANT: DSM N.V. 5 <120> TIŢĻE OF INVENTION: Genes encoding enzymes in the biosynthetis of pimaricin and applications 6 thereof 7 <130> FILE REFERENCE: S-2959 9 <140> CURRENT APPLICATION NUMBER: 10/018,009A C--> 10 <141> CURRENT FILING DATE: 2001-12-14 12 <160> NUMBER OF SEQ ID NOS: 14 14 <170> SOFTWARE: PatentIn Ver. 2.1 ENTERE 16 <210> SEQ ID NO: 1 17 <211> LENGTH: 5544 18 <212> TYPE: DNA 19 <213> ORGANISM: Streptomyces natalensis 21 <220> FEATURE: 22 <221> NAME/KEY: CDS 23 <222> LOCATION: (1)..(5544) 25 <400> SEQUENCE: 1 26 atg gtg ccc gtc cac aca gat gac tac gcg atc cag ccg ccc gcc gac 27 Met Val Pro Val His Thr Asp Asp Tyr Ala Ile Gln Pro Pro Ala Asp 10 30 acc gcg cat gga ggg ggt ggc ttc aca ctg cct gct gtg ttc gag gcc 31 Thr Ala His Gly Gly Gly Phe Thr Leu Pro Ala Val Phe Glu Ala 32 20 34 gcc gtg gaa tcc gcc ccc gac gca gtg gcc ctc gtc gac ggc acc gtt 35 Ala Val Glu Ser Ala Pro Asp Ala Val Ala Leu Val Asp Gly Thr Val 40 38 cet gga cet ggg ege atg tgg ege gea gae gte gat geg etg gee ege 39 Pro Gly Pro Gly Arg Met Trp Arg Ala Asp Val Asp Ala Leu Ala Arg 55 42 ggg ttg cag gag tcg ggt atc gca ccc ggt gac gtg gtg gcg gtg cgg 43 Gly Leu Gln Glu Ser Gly Ile Ala Pro Gly Asp Val Val Ala Val Arg 46 ctg ccg aac tgc ggg agg ttt ccc acc ctg cac ctg gcc gtc gcg gcc 288 47 Leu Pro Asn Cys Gly Arg Phe Pro Thr Leu His Leu Ala Val Ala Ala 50 gtc gga gcg gta ctg ctg ccc atc cac cag ggc acc ccg ctc ccg gag 336 51 Val Gly Ala Val Leu Leu Pro Ile His Gln Gly Thr Pro Leu Pro Glu 52 100 54 gtc gac gcg ctg ctg acc cgg gcg gaa ccc gcc ctt ctc gtc ctg tcg 55 Val Asp Ala Leu Leu Thr Arg Ala Glu Pro Ala Leu Leu Val Leu Ser 120 58 gcc gcc ggg agc gac ggc ctg gcg acg gcc cgt tcg ctt ctg gag agc 432 59 Ala Ala Gly Ser Asp Gly Leu Ala Thr Ala Arg Ser Leu Leu Glu Ser

135

140

130

Input Set : N:\FATIMA\10018009.txt

63				ctg Leu													480
				tcc													528
67 68	Glu	Ser	Gly	Ser	Val 165	Gly	Gly	Gly	Glu	Ser 170	Gly	Ser	Gly	Arg	Arg	Ser	
	ctg	gac	999	ctg		gcc	ggc	tgg	gcg		agc	999	ccc	cgg		gtg	576
	Leu	Asp	Gly	Leu	Leu	Ala	Gly	Trp		Gly	Ser	Gly	Pro	_	Pro	Val	
72	G2G	ata	200	180	ast	ata	~~~	ata	185	a+ a	at a	000	t aa	190	~~~		624
				ccc Pro													624
76	11.55	• • • • • • • • • • • • • • • • • • • •	195	110	пор	1100	110	200	val	<u> </u>	Val	110	205	DCI	Cly	1111	
78	gtc	tcg	gca	cgg	ccc	aaa	ctg	tgt	gtg	cac	agc	cac	gac	ggt	ctg	ctg	672
	Val		Ala	Arg	Pro	Lys		Cys	Val	His	Ser		Asp	Gly	Leu	Leu	
80		210					215					220					=
				gca Ala													720
	225	ASII	1111	ATA	AIA	230	1111	Ата	GIU	Ата	235	Asp	AIA	FIIE	ASP	240	
		gtc	ctc	acg	qcc		ccq	atq	acc	cat		ttc	qqc	ctc	caq		768
				Thr													
88					245					250			_		255		
				gcg													816
	Leu	His	Ala	Ala	Leu	Phe	Ala	Ala		Thr	Gln	Val	Leu		Thr	Gly	
92	.			260					265					270			0.5.4
				gac Asp													864
96	пр	Asp	275	Asp	Arg	FIIE	ьец	280	GIII	Ala	Arg	GIU	285	GIY	PIO	Arg	
	atc	ata	_	gcc	ata	ccc	acc		cta	caa	gac	atc		aca	cga	cta	912
				Ala													
10		290					295			_	_	300			_		
																acc	960
			g Thi	r Gly	/ Gli			. Gly	r Phe	Thr		_	Glr	ı Val	L Arg	g Thr	
	4 305					310					319					320	
																c gtc a Val	1008
10		a GI	, TT	a Ala	325		PIC	ATC	шес	330		r Ard	y vai	r WI	33!		
		gad	c tac	gaa			ato	ato	ı tac			a tco	gad	ato		acc	1056
																y Thr	
11:		_	_	340					345	_				350	_	•	
																g ccg	1104
		Thi			Arg	, Ala	His			Asp	Gly	у Суя			/ Glu	ı Pro	
110	-		355					360					365				
																tgc	1152
120		370		, val	. ASI	, val	375		. val	. ASP	, GI	380	_	GII	ı GIL	ı Cys	
				gaq	aac	aat			cac	tac	cac			a aac	tte	ttc	1200
																ı Phe	
	385		•		-	390				4	395			•		400	
126	cgc	gg	tat	ttc	cgc	gaa	ccg	gag	cts	g acc	g	teg	gco	cto	acc	gac	1248

Input Set : N:\FATIMA\10018009.txt

127 128	Arg	Gly	Tyr	Phe	Arg 405	Glu	Pro	Glu	Leu	Thr 410	Arg	Ser	Ala	Leu	Thr 415	Asp	
130	qac	qqc	tgg	ctq	cqq	acc	ggc	gat	ctc	qcq	acc	qtc	qac	qcq	qac	qqc	1296
						Thr											
132	-	-	-	420			_	_	425				-	430	-		
134	qtq	qtq	qtq	ctq	cac	ggc	cqq	qcq	qcc	gag	ctq	atc	aat	acq	qqq	qqc	1344
						Gly											
136			435			-	_	440					445		•	-	
138	cqq	aaq	ttc	aqc	gcc	ggc	gaq	qtc	gag	qqa	ctg	ctc	tcc	qqc	ttt	acq	1392
		_		_	_	Gly		_			_					_	
140	_	450				-	455			•		460		•			
142	gac	ctg	ggc	ccg	ctg	gcc	gtc	gtc	ggc	gca	ccg	gac	gac	cgg	ctc	ggg	1440
						Āla											
144	465		_			470			_		475	_	_	_		480	
146	gag	tac	ccg	tgc	ctg	gtc	gtg	acc	gac	cac	gcg	gac	ggc	acc	atc	ggc	1488
						Val											
148		_		_	485				_	490		_	_		495	_	
150	ctg	agc	gag	gtg	acc	gcg	ttc	ctg	cgc	cgg	ctg	ggg	ctc	gcc	gac	cac	1536
151	Leu	Ser	Glu	Val	Thr	Ala	Phe.	Leu	Arg	Arg	Leu	Gly	Leu	Ala	Asp	His	
152				500					505					510			
154	aag	atc	CCC	ctc	gaa	ctg	gtc	acc	gtg	cgc	gag	ctt	ccc	ttc	tcc	cct	1584
155	Lys	Ile	${\tt Pro}$	Leu	Glu	Leu	Val	Thr	٧al	Arg	Glu	Leu	${\tt Pro}$	Phe	Ser	Pro	
156			515					520					525				
158	gcc	ggg	aag	ctc	gac	cgc	999	gcg	ctc	aag	cgg	ctg	ctc	gcc	aac	ctt	1632
159	Ala	Gly	Lys	Leu	Asp	Arg	Gly	Ala	Leu	Lys	Arg	Leu	Leu	Ala	Asn	Leu	
160		530					535					540					
						ccg											1680
		Glu	Val	Ser	Val	Pro	Ala	Arg	Leu	Gly		Val	Pro	Pro	Tyr		
	545					550		`			555					560	
						gac											1728
	Ala	Glu	Glu	Ala		Asp	Leu	Val	Arg	_	Cys	Val	Gly	Arg		Leu	
168					565					570					575		,
						gcc											1776
	Arg	Tyr	GLY	_	Ala	Ala	Val	Pro		Pro	Pro	Asp	Lys	_	Phe	Phe	
172				580					585					590			1004
		_	-	_	-	ttc	_	_			_	-					1824
	ser	Pro	_	гàг	Asp	Phe	Arg		ьeu	GIY	ьeu	Asp		тте	GIĀ	Ala	•
176	~+~	~~~	595			~+~	~-~	600	~~~	~~~		~~~	605	~~~	a+ a		1072
						ctg											1872
180	val		ьец	Arg	ASII	Leu	615	Arg	GIU	GIU	THE	620	ьeu	PIO	ьец	PIO	
	~~~	610	ata	~~~	++-	gac			2.at	222	~~~		at a	~~~	~~~	ata	1920
						Asp											1920
	625	TILL	пеп	AIG	FIIE	630	Ser	FIO	1117	FIO	635	MIG	vaı	AId	Ary	640	
		acc	asa	020	asa	gag	ccc	taa	020	aea.		cca	200	asa	220		1968
	_	_		_		Glu	_		_	_		_				_	1,00
188	LCU	nia	JIU	3111	645	Jiu	110	DGI	3111	650	GIU	110	y	JIU	655	210	
	aca	gac	aat	acc		ccg	ata	aca	atc		aac	ato	acc	tac		cta	2016
						Pro											2010
			I		1.50			<del>-</del> u		• 41				<b>-</b> 13	9	u	

Input Set : N:\FATIMA\10018009.txt

192				660					665					670			
	cca	aac	aaa		asc	tcc	ccc	gac		cta	taa	mam	cta		~~~	<b>~</b> 3 <i>~</i>	2064
								Asp									2064
	PIO	Gry	675	мта	Asp	Ser	PIO		Ald	Leu	пр	Gru		ьeu	Ala	Asp	
196								680					685				
								ttc									2112
	GIŸ		Asp	Ala	Met	ser		Phe	Pro	Thr	Asp	-	Gly	Trp	Asp	Leu	
200		690					695					700					
								gcc									2160
203	Asp	Arg	Leu	Phe	Asp	Glu	Asp	Ala	Asp	Arg	Pro	Gly	Thr	Ser	Tyr	Ala	
	705					710					715					720	
								gac									2208
207	Arg	Glu	Gly	Gly	Phe	Leu	His	Asp	Ala	Gly	Asp	Phe	Asp	Ala	Gly	Phe	
208					725					730					735		
210	ttc	ggc	ctg	tcg	gac	cag	gag	gcg	acg	gcg	acc	gat	ccg	cag	cag	cgg	2256
								Ala									
212		•		740	•				745			_		750		- 3	
214	cta	ctt	cta	gag	aca	acc	taa	gag	acc	ttc	gag	caa	aca		atc	gac	2304
								Glu									
216			755					760			014	****9	765	0-1	110	1101	
	cca	cac		cta	add	aaa	add	cgt	200	aac	ata	ttc		aaa	aca	ata	2352
								Arg									2332
220	110	770	DCI	пец	Arg	Gry	775	A. 9	1111	GIY	vai	780	1111	GIY	Ата	Mec	•
	~~~		~~~	+ - +	~~~	200		gcg	+	~~~	~~~			~~~		~~~	2400
																	2400
		Arg	GIY	ıyı	GIY		ser	Ala	ser	AId		PIO	ser	Ala	ттр		
	785					790					795				_	800	
226	agc	atg	CEC	atc	acc	999	acc	gcc	ggc	agc	gcg	gtc	tcg	999	cgc	atc	2448
	ser	мес	Leu	тте		GIY	Thr	Ala	GIA		Ala	vai	ser	GLY		IIe	
228					805					810					815		
								ggc									2496
	Ala	Tyr	Thr		GLY	Leu	Glu	Gly		Ala	Leu	Thr	Val	_	Thr	Ala	
232				820					825					830			
								ctg									2544
	Ser	Ser		Ser	Leu	Val	Ala	Leu	His	Leu	Ala	Cys	Arg	Ser	Leu	Arg	
236			835					840					845				
								ctg									2592
239	Ser	Gly	Glu	Thr	Asp	Leu	Ala	Leu	Ala	Gly	Gly	Val	Thr	Val	Met	Ala	
240		850					855					860					
242	acc	ccg	gcg	CCC	ttc	gca	cac	ttc	tcc	cgg	ctg	cgc	gcg	ctg	tcc	CCC	2640
243	Thr	Pro	Ala	${\tt Pro}$	Phe	Ala	His	Phe	Ser	Arg	Leu	Arg	Ala	Leu	Ser	Pro	
244	865					870					875					880	
246	gac	tcc	cgc	tcc	atg	gcg	tac	gcg	gac	gcc	gcg	aac	ggc	tcg	gcg	tgg	2688
								Ala									
248	_		_		885		-		-	890			•		895	-	
250	tca	qaq	qqc	qca	qaa	cta	ctt	ctg	cta		caa	cta	agc	qac		caa	2736
								Leu									
252			- 2	900	2				905					910		9	
	cac	aac	gga		cat	atc	cta	gcg		ata	caa	ממר	t.c.c		ata	aat	2784
								Ala									2,03
256	3		915		9	. ~ .		920	 _u	*41	9	- y	925			- 111	
								220					723				

Input Set : N:\FATIMA\10018009.txt

							999 Gly										2832
260		930	_				935					940	_				
							gcc										2880
263	Gln	Arg	Val	Ile	Arg	Gln	Ala	Leu	Ala	Asp	Ala	Gly	Leu	Thr	Pro	Gln	
264	945					950					955					960	
266	gac	gtg	gac	gcc	gtg	gag	ggg	cac	ggc	acc	ggc	acg	ccg	ctc	ggc	gac	2928
							Gly										
268	•		-		965		-		•	970	•				975	•	
270	ccc	atc	gag	aca	cag	aca	ctg	cta	acc	acq	tac	aac	caa	caq	caa	cct	2976
							Leu										
272				980					985		-1-	1		990			
	ata	gaa	caa		tta	taa	ctg	aaa		ata	aad	tra	aac		.aaa	cac	3024
		_		_	_		Leu		_		_	_					3021
276	Val	Olu	995	110	пси	110	LCu	1000		vai	Lys	JCI	1005		Ory	1110	
_	202	caa		acc	acc	aaa	gtc			ata	atc	220			ata	aca	3072
							Val										3072
280	1111	1010		AIG	HIU	OLY	1015		Cry	Vai	110	1020		Val	neu	ALG	
	cta			aaa	ata	cta	ccg		200	ata	Cac			act	000	taa	3120
							Pro										3120
	1025		птэ	Gry	vai	1030		GIII	1111	пеп	1039		Asp	Ата	PIO	1040	
			~+~	~~~				~~+							~~~		2160
							gcc										3168
	Ата	ьys	vaı	Asp	_		Ala	GIY	ser		_	ьeu	ьeu	Thr			
288					1045				_	1050					1055		
							agc										3216
	Arg	Pro	Trp		_	GIU	Ser	GIY	_		Arg	Arg	Ата	_		ser	
292				1060					1069					1070			
							acc										3264
	Ser	Phe			Thr	GIY	Thr			Hıs	Val	He			GIu	Ala	
296			1075					1080					1089				
							999			_		_					3312
	Pro	_		Ala	Ala	Ala	Gly		Arg	Ala	Glu			Glu	Glu	Ala	
300		1090					1099					1100					
							gct										3360
303	Arg	Cys	Ala	Ser	Ser	Pro	Ala	Arg	Leu	Pro	Glu	Pro	Pro	Gly	Asp		
	1109					1110					1115					1120	
							tcc										3408
307	Ala	Ala	Pro	Trp	Val	Leu	Ser	Ala	Arg	Ser	Arg	Ala	Ala	Leu	Arg	Ala	
308					1125	5				1130)				1135	5	
							gac										3456
311	Gln	Ala	Leu	Arg	Leu	Ala	Asp	Gln	Val	Ala	Ala	Asp	Pro	Gly	Leu	Arg	
312				1140)				1145	5				1150)		
314	gcc	cag	gat	gtc	gcc	cat	gcc	ctg	gcc	acc	tcc	cgt	acc	ctg	cac	cgg	3504
315	Ala	Gln	Asp	Val	Ala	His	Ala	Leu	Ala	Thr	Ser	Arg	Thr	Leu	His	Arg	
316			1155	5				1160)				1165	5			
318	cac	cgg	gcc	gtc	gtc	agc	ggc	tcc	gac	cgg	gca	caa	atg	ctc	gcc	gcg	3552
							Gly										
320		1170					1175		-	_		1180					
	gca	aaq	cgq	ttc	ggg	ctc	ggt	gag	cgq	acc	gcq	ggc	gtc	acc	ccq	gac	3600
	-	_											_		_	_	

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/018,009A

DATE: 08/16/2004 TIME: 17:10:34

Input Set : N:\FATIMA\10018009.txt

Output Set: N:\CRF4\08162004\J018009A.raw

L:7 M:283 W: Missing Blank Line separator, <130> field identifier L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date